# *Democracy's Destruction?*
# *Changing Perceptions of the Supreme Court, the Presidency, and the Senate after the 2020 Election*

***James L. Gibson***

Version 8

## ONLINE APPENDIX

### APPENDIX 2.1: MEASURES AND MEASUREMENT THEORY

Because Chapter 2 addresses important issues of the conceptualization and measurement of closely related concepts, I allocate space in this appendix to measurement issues and results than I might not otherwise. I begin with some general theories and principles of measurement.

#### Measurement Theory

Multiple indicator scales have great advantages over single-item indicators, not the least reason of which is that the former are readily susceptible to assessments of validity and reliability whereas the latter are usually not. Validity and reliability are, of course, the two main indicia of the quality of measurement. When embedded within a measurement theory such as Classical Test Theory (CTT), which posits a latent construct representing the concept, the advantages of this approach multiply. Finally, Common Factor Analysis (CFA) is a psychometric technique that aligns extremely well with Classical Test Theory.

CTT posits that the variance in any given indicator can be decomposed into two sources: variance associated with the concept and variance unique to the indicator. Conceptually, validity is the degree to which the indicator is correlated with the concept. Empirically, validity is the degree to which the indicator is correlated with the empirical indicator of the latent construct. By

definition, variance not associated with the latent construct is unique to the indicator. Unique variance can be further decomposed into random variance—a threat to reliability—and systematic variance—a possible threat to validity. In causal measurement diagrams, this is why the indicator is represented as a dependent variable: its observed variance is caused by the latent construct and by unique factors.

One implication of this theory is that any given indicator should not be treated as a perfect measure of the latent construct. The latent construct is the best empirical representation of the concept, not individual items. Without getting too far ahead of ourselves, the correlation between the item and the latent construct is typically represented by a factor loading: a bivariate correlation, which may be thought of as a validity coefficient. Moreover, individual items in a scale are likely for various reasons to have different validity coefficients.

Because an indicator does not perfectly represent the construct, one must be a bit careful about using face validity as a criterion for assessing the quality of measurement. An indicator may, on its face, not be a perfect measure of the concept. But empirical analysis, under the assumptions of Classical Test Theory, can partition observed variance into that associated with the latent construct and that not associated. I do not, of course, endorse the use of shoddy indicators, but the degree to which a particular item is not well connected to the concept can be empirically assessed and corrected for.[1]

[1] Gibson (1994, 99-100) has shown how items with very weak validity can be identified and corrected for. For instance, in surveys in the early days of the transition of the Soviet Union, one of the measures of dogmatism did not work—it did not measure the concept it was intended to measure. Similarly, an Inglehart measure of democratic political culture was found to have

Ideally, a set of indicators would vary widely in their "strength"—that is, how extreme the item is, how much support is necessary to endorse the question. The diffuse support item about doing away with the Supreme Court does not require very much court support to reject (at least in the U.S.). Those with unlimited loyalty would not endorse the item, just as most of those with only tepid loyalty would not endorse the item. In the American context, a measure with a skewed distribution fails to offer much power to discriminate among respondents; however, the same item that is skewed among American respondents may not be anywhere nearly as skewed in other national contexts (see Gibson, Caldeira, and Baird 1998). To the extent possible, items in a multiple-item set should range from roughly 20/80% (agree/disagree, on a dichotomized measure) to 80/20%. Constructing meaningful items with this degree of variability is often fairly challenging, however.

Items that simply parrot each other may seem to have strong psychometric properties (i.e., they are strongly intercorrelated), but may fail to offer much discriminatory power.[2] Thus, varying the substantive content of an item-set is desirable. In the end, the degree to which any given item is a valid measure of a concept (and validity is indeed a continuum) can be assessed empirically through Common Factor Analysis.

The simplest measurement hypothesis that one can test is that an item-set is unidimensional; that is, the variance in the set can be adequately represented by a single

---

very low validity owing to the different ways that "revolutionary change" is understood in different cultural contexts (e.g., the US and USSR).

[2] This is one reason why very high Cronbach's alpha scores are actually *undesirable*.

underlying latent construct.[3] In some instances, this hypothesis must be rejected.[4] The degree to which multiple sub-dimensions are correlated is an empirical question and should not be assumed. Moreover, an assumption of orthogonality in particular is very rarely theoretically justified.

Common Factor Analysis (also known as principal factor analysis (PFA) or principal axis factoring (PAF)) fits Classical Test Theory extremely well. Because dimensionality can easily become fairly complicated and to some degree arbitrary (see studies of the dimensionality of voting in Congress), the simplest and safest approach is to begin with the hypothesis that an item-set is unidimensional. Especially if the items were designed to be unidimensional, this hypothesis is quite reasonable. The test of the unidimensionality hypothesis is essentially a confirmatory test of whether factors beyond the first factor extracted are sufficiently powerful to

---

[3] Multi-dimensional structures are more difficult to analyze, in part because one must make some assumptions about the degree to which the sub-dimensions are intercorrelated. Often, the assumption that the sub-dimensions are orthogonal to one another is made; this is almost always the least defensible assumption.

[4] For instance, Gibson and Duch (1992a, 1992b) sought to measure anti-Semitism in the Former Soviet Union with a set of propositions. However, empirical analysis demonstrated that the item-set required two dimensions to understand effectively. At least in the Former Soviet Union, it seems that anti-Semitism can fruitfully be thought of as reflecting a) stereotyping and prejudice toward Jews, and b) support for discrimination against Jews. In that case, these two dimensions were certainly correlated, but not so well correlated that they could be collapsed into a single dimension.

warrant attention.[5] "Sufficiently powerful" can be operationalized in a number of ways (e.g., so-called scree tests), but it is conventional to determine whether factors beyond the first factor extracted are associated with an eigenvalue greater than 1.0. If only a single factor has an eigenvalue greater than 1.0, then unidimensionality is established.

Common Factor Analysis is named as it is in part because it only focuses on "common inter-item variance" not all inter-item variance. The implication of this is that the commonalities (the diagonal) in the correlation matrix are not 1.0; the assumption that an item is perfectly correlated with itself is *not* imposed. Instead, the assumption is made that only that variance shared between an item and the entire set of other items is eligible for analysis; hence, the term "common variance." Because every measure must include some random error, it makes little sense to assume that even the repeated measures of exactly the same item will generate observed correlations of 1.0. The diagonal in the inter-item matrix in CFA therefore often substitutes the squared multiple correlation coefficient for 1.0, acknowledging the lack of perfect reliability in the measures.

The most important conclusion to be drawn from this exegesis is that it is important to assess the validity and reliability of operationalizations of concepts, and Classical Test Theory connected to Common Factor Analysis is an excellent way to do so.

I turn now to the measures of diffuse support.

[5] In this sense, the distinction between exploratory and confirmatory factor analysis is often blurred.

**Measuring Diffuse Support for the U.S. Supreme Court**

The first latent construct I focus on in this appendix is institutional support for the U.S. Supreme Court. My thinking about operationalizing institutional support (or loyalty) follows a considerable body of research on theorizing about and measuring mass perceptions of high courts (see Caldeira and Gibson 1992, Gibson, Caldeira, and Baird 1998, Caldeira and Gibson 1995, Gibson 2007, and Gibson and Caldeira 1992, 1995, 1998, 2003, 2009).[6] That research conceptualizes loyalty as opposition to making fundamental structural and functional changes in the institution (see Boynton and Loewenberg 1973), and is grounded in the history of attacks by politicians against courts in the U.S. (see Caldeira 1987) and elsewhere (e.g., manipulation of their jurisdiction — see Schwartz, Behrens, and Lorber 2000). As Caldeira and Gibson describe it (1992, 638), those who have no loyalty toward the U.S. Supreme Court are willing "to accept, make, or countenance major changes in the fundamental attributes of how the high bench functions or fits into the U.S. constitutional system" (see also Loewenberg 1971). To the extent that people support fundamental structural changes in an institution, they are extending little legitimacy to that institution. Conceptually, loyalty thus ranges from complete unwillingness to support the continued structure and function of the institution to staunch institutional fealty. Consequently, my measure of the legitimacy of the Supreme Court is derived from that used by Gibson, Caldeira, and Spence (2003) and Gibson and Caldeira (2009).

[6] For a full explication of the conceptual and theoretical meaning of this concept see the discussion in Caldeira and Gibson (1992, 636-642). Here, I provide only an overview of the conceptualization since this is well-trodden territory.

Table 2.1.1 reports the frequency distributions of the responses to these statements in the July 2020 survey. Note that responses to these items were collected via a five-point Likert response set with uncertainty ("neither agree nor disagree") at the center-point of the scale. Volunteered "don't know" responses (which were quite rare) were recoded to uncertainty, although those who failed to answer the question were treated as missing data.

What conclusions should one draw, therefore, about whether the U.S. Supreme Court enjoys a "reservoir of goodwill" among the American people? On the one hand, a plurality of the American people—and a majority of those offering a substantive opinion—do not want to do away with the Supreme Court, do not support restricting the Court's jurisdiction, and do not wish to remove judges because of how they vote. On the other hand, a plurality sees the Court as getting too mixed up in politics, and significant support exists for making the Court less independent. And, just as the American people favor electing state court judges (Gibson 2012; Geyh 2019), they tend to favor more accountability from the Supreme Court. In these data, the "supportive" replies are clearly significantly more common than "not supportive" replies on three of the six items. On two of the other three, non-supportive replies are somewhat more common, and on one item support and non-support are roughly equally common. On this set of proposals, the Supreme Court support glass is "clearly" either half full or half empty!

**Table 2.1.1. Diffuse Support for the United States Supreme Court, July 2020**

| | Level of Diffuse Support for the Supreme Court | | | | | |
|---|---|---|---|---|---|---|
| | Percentage | | | | | |
| Indicator | Not Supportive of the Institution | Undecided | Supportive of the Institution | Mean | Std. Dev. | N |
| Do away with the Court | 13.8 | 26.1 | 60.1 | 3.6 | 1.1 | 940 |
| Restrict Court's Jurisdiction | 20.0 | 34.0 | 46.0 | 3.4 | 1.0 | 943 |
| Too mixed up in politics | 37.0 | 42.1 | 20.9 | 2.8 | 1.0 | 942 |
| Remove judges who rule against majority | 21.3 | 40.3 | 38.4 | 3.3 | 1.1 | 939 |
| Makes Court less independent | 32.4 | 33.5 | 34.1 | 3.1 | 1.2 | 945 |
| Control the actions of the Court | 31.8 | 37.3 | 30.9 | 3.0 | 1.1 | 943 |

Note: The percentages are calculated on the basis of collapsing the five-point Likert response set (e.g., "agree strongly" and "agree" responses are combined) and sum to 100% across the three percentage columns (except for rounding errors). The percentage "Supportive of the Institution" is the percentage of respondents giving a reply supportive *of the Court*, not of the statement itself. The means and standard deviations are calculated on the uncollapsed distributions. Higher mean scores indicate more institutional loyalty.

The propositions (asked with a five-point Likert response set) are:

Do away with the Court: If the U.S. Supreme Court started making a lot of decisions that most people disagree with, it might be better to do away with the Court altogether.

Restrict Court's Jurisdiction: The right of the U.S. Supreme Court to decide certain types of controversial issues should be reduced.

Too mixed up in politics: The U.S. Supreme Court gets too mixed up in politics.

Remove judges who rule against majority: Judges on the U.S. Supreme Court who consistently make decisions at odds with what a majority of the people want should be removed from their position as judge.

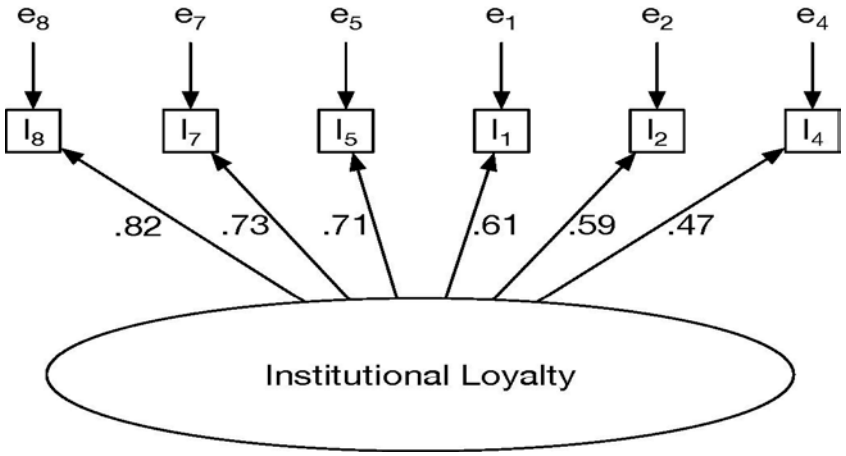
Makes Court less independent: The U.S. Supreme Court ought to be made less independent so that it listens a lot more to what the people want.

Control the actions of the Court: It is inevitable that the U.S. Supreme Court gets mixed up in politics; therefore, we ought to have stronger means of controlling the actions of the U.S. Supreme Court.

Source: NORC AmeriSpeak, July 2020.

Figure 2.1.1 reports the results of the factor analysis of this six-item set. Institutional loyalty is the latent construct; the factor loadings reported in the figure can be understood as validity coefficients (the correlation of the indicator and the construct). The most valid measure of diffuse support is the item about controlling the Court; the least valid measure is the assertion that the Court gets too mixed up in politics. Still, the empirical results indicate that these are all useful indicators of the concept, although to varying degrees.

**Figure 2.1.1. Measurement Model, Diffuse Support for the U.S. Supreme Court**




The propositions (asked with a five-point Likert response set) are:

> I1. If the U.S. Supreme Court started making a lot of decisions that most people disagree with, it might be better to do away with the Court altogether.
> I2. The right of the U.S. Supreme Court to decide certain types of controversial issues should be reduced.
> I4. The U.S. Supreme Court gets too mixed up in politics.
> I5. Judges on the U.S. Supreme Court who consistently make decisions at odds with what a majority of the people want should be removed from their position as judge.
> I7. The U.S. Supreme Court ought to be made less independent so that it listens a lot more to what the people want.
> I8. It is inevitable that the U.S. Supreme Court gets mixed up in politics; therefore, we ought to have stronger means of controlling the actions of the U.S. Supreme Court.

The analysis in this project relies upon a six-item measure of diffuse support. The set of diffuse support indicators has quite strong psychometric properties. Reliability is high—Cronbach's alpha = .78. So too is validity. The item set is strongly unidimensional (the second eigenvalue from a Common Factor Analysis (CFA) is a mere .75), and all items load well on the first unrotated factor (minimum loading = .47). Because a summated index is very strongly correlated with the factor score from the CFA (r = .97), I use that index in my analyses. These measurement results clearly indicate that this summary measure is quite strong in terms of both validity and reliability.

## REFERENCES, ONLINE APPENDIX


Boynton, G. R. and Gerhard Loewenberg. 1973. "The Development of Public Support for Parliament in Germany 1951-59." *British Journal of Political Science* 3: 169-89.

Caldeira, Gregory. 1987. "Public Opinion and the American Supreme Court: FDR's Court-Packing Plan." *American Political Science Review* 81 (#4):1139-1153.

Caldeira, Gregory A., and James L. Gibson. 1992. "The Etiology of Public Support for the Supreme Court." *American Journal of Political Science* 36 (#3, August): 635-664.

Caldeira, Gregory A., and James L. Gibson. 1995. "The Legitimacy of the Court of Justice in the European Union: Models of Institutional Support." *American Political Science Review* 89 (#2, June): 356-376.

Geyh, Charles. 2019. *Who is to Judge: The Perennial Debate Over Whether to Elect or Appoint America's Judges*. New York, NY: Oxford University Press.

Gibson, James L. 1994. "Survey Research in the Past and Future USSR: Reflections on the Methodology of Mass Opinion Surveys." In *Research in Micropolitics: New Directions in Political Psychology, Volume 4*. Edited by Michael X. Delli Carpini, Leonie Huddy, and Robert Y. Shapiro. Greenwich, CT: JAI Press Inc. Pp. 87-114.

Gibson, James L. 2007. "The Legitimacy of the U.S. Supreme Court in a Polarized Polity." *Journal of Empirical Legal Studies* 4 (#3, November): 507-538.

Gibson, James L. 2012. *Electing Judges: The Surprising Effects of Campaigning on Judicial Legitimacy.* Chicago: University of Chicago Press.

Gibson, James L., and Gregory A. Caldeira. 1992. "Blacks and the United States Supreme Court: Models of Diffuse Support." *Journal of Politics* 54 (#4, November): 1120-1145.

Gibson, James L., and Gregory A. Caldeira. 1995. "The Legitimacy of Transnational Legal Institutions: Compliance, Support, and the European Court of Justice." *American Journal*

*of Political Science* 39 (#2, May): 459-489.
Gibson, James L., and Gregory A. Caldeira. 1998. “Changes in the Legitimacy of the European Court of Justice: A Post-Maastricht Analysis.” *British Journal of Political Science* 28 (#1, January): 63-91.
Gibson, James L., and Gregory A. Caldeira. 2003. “Defenders of Democracy? Legitimacy, Popular Acceptance, and the South African Constitutional Court.” *The Journal of Politics* 65 (#1, February): 1-30.
Gibson, James L., and Gregory A. Caldeira. 2009. *Citizens, Courts, and Confirmations: Positivity Theory and the Judgments of the American People*. Princeton, NJ: Princeton University Press.
Gibson, James L., Gregory A. Caldeira, and Vanessa Baird. 1998. “On the Legitimacy of National High Courts.” *American Political Science Review* 92 (#2, June): 343-358.
Gibson, James L., Gregory A. Caldeira, and Lester Kenyatta Spence. 2003. “Measuring Attitudes toward the United States Supreme Court.” *American Journal of Political Science* 47 (#2, April): 354-367.
Gibson, James L., and Raymond M. Duch. 1992a. “Attitudes Toward Jews and the Soviet Political Culture.” *Journal of the Soviet Nationalities* 2 (Spring): 77-117.
Gibson, James L., and Raymond M. Duch. 1992b. “Anti-Semitic Attitudes of the Mass Public: Estimates and Explanations Based on a Survey of the Moscow Oblast.” *Public Opinion Quarterly* 56 (#1, 1992): 1-28.
Loewenberg, Gerhard. 1971. "The Influence of Parliamentary Behavior on Regime Stability." *Comparative Politics* 3: 177-200.
Schwartz, Victor E., Mark A. Behrens, and Leah Lorber. 2000. “Federal Courts Should Decide Interstate Class Actions: A Call for Federal Class Action Diversity Jurisdiction Reform.” *Harvard Journal on Legislation* 37: 483-514.